

Concept Formation

Robert L. Goldstone

Thomas T. Hills

Samuel B. Day

Indiana University

Correspondence Address: Robert Goldstone  
Department of Psychology  
Indiana University  
Bloomington, IN. 47408

Other Correspondences: [rgoldsto@indiana.edu](mailto:rgoldsto@indiana.edu)  
(812) 855-4853

Keywords: Concepts, Learning, Representation, Prototypes, Exemplars

## Concept Formation

A concept is a mentally possessed idea or notion that can be used to categorize information or objects. Over the course of each person's lifetime, thousands of concepts are learned, for nouns like **corkscrew**, **justice**, and **doorknob**, adjectives like **green**, **symmetric**, and **beautiful**, and verbs like **kick**, **climb**, and **eschew**. While some philosophers have maintained that we do not genuinely learn new concepts through induction (Fodor, 1988), most psychologists believe that concepts can be learned, and that the representational capacity of the learner increases as they acquire new concepts. Most efforts have been spent developing accounts of how people acquire and represent concepts, including models based on: **rules**, **prototypes**, **exemplars**, **boundaries**, and **theories**.

According to rule-based models of concept formation, to acquire a concept is to learn the rule that allows one to determine whether or not the concept applies to an entity. The most influential rule-based approach to concepts may be Bruner, Goodnow, and Austin's (1956) hypothesis testing approach. In a typical experiment, their subjects were shown flash cards that had different shapes, colors, quantities, and borders. The subjects' task was to discover the rule for categorizing the flash cards by selecting cards to be tested and by receiving feedback from the experimenter indicating whether the selected card fit the categorizing rule or not. However, at least three criticisms have been levied against this rule-based approach. First, it has proven to be very difficult to specify the defining rules for most concepts. Even a seemingly well-defined concept such as *bachelor* seems to involve more than its simple definition of "Unmarried male." The counter-example of a five-year old child (who does not really seem to be a bachelor) may be fixed by adding in an "adult" precondition, but an indefinite number of other preconditions are required to exclude a man in a long-term but unmarried relationship, the pope, and a 80-year old widower with 4 grandchildren. Second, the category membership for many objects is not clear. People disagree with each other, and even themselves from one occasion to the next, on whether or not a starfish is a fish, a camel is a vehicle, a hammer is a weapon, and a stroke is a disease. Third, even when a person shows

consistency in placing objects in a category, people do not treat the objects as equally good instances of the concept. When asked to rate the typicality of animals like robin and eagle for the concept *bird*, or chair and hammock for the concept *furniture*, subjects reliably give different typicality ratings for different objects

In response to these criticisms, the prototype model of concept formation, was developed. Central to Eleanor Rosch's development of prototype theory is the notion that concepts are organized around family resemblances rather than rules (Rosch, 1975). The prototype for a concept consists of the most common attribute or shared feature values associated with the instances of the concept. Once prototypes for a set of concepts have been determined, categorizations can be predicted by determining how similar an object is to each of the prototypes. An empirical result consistent with the prototype view is that never-before-seen prototypes are often categorized as accurately as previously seen distortions of the prototype, and with higher accuracy than new distortions. One interpretation of these results is that prototypes are extracted from resemblances among distortions, and used as a basis for determining categorizations.

The third approach, exemplar modeling, denies that prototypes are explicitly extracted from individual cases. Instead, in exemplar models, to learn a concept is simply to store the actual individual cases that have been observed. An exemplar model forms the concept **bird** by representing all of the instances (exemplars) that belong to this category, including robins, eagles, and penguins. The exemplar approach assumes that categorization decisions are based on the similarity of the object to be categorized to all of the exemplars of each relevant concept (Nosofsky, 1986). Consistent with an exemplar account, evidence suggests that people show good categorization of new stimuli just to the extent that the new stimuli resemble previously learned members of a category. Instead of viewing rote memorization of exemplars as evidence of human stupidity or laziness, exemplar accounts argue that preserving the entire stimulus in its full richness of detail provides the most flexibility for future categorization and reasoning.

A fourth approach claims that a concept is formed when its category boundary has been established (Ashby & Townsend, 1986). Whereas the prototype model represents a concept by its

center--its most typical member--a category boundary model represents the concept by its perimeter. Frequently, category boundary models are constrained by assuming that the boundary that a person constructs to characterize a concept is the one that maximizes the likelihood of correctly categorizing an object into the concept. To learn a categorization is to learn the parameters that define the lines or curve that act as the boundaries between two or more categories. One problem with this account is that the boundary between two categories depends on several contextual factors that make some prototypes or exemplars more influential. In many cases, it is more parsimonious to hypothesize representations for the concept considered by itself, and view category boundaries as side effects of the competition between neighboring categories.

The fifth approach, theory-based concept formation, begins with the observation that the four preceding approaches all operate without regard to the actual meaning of the concepts. By contrast, many researchers have argued that the use of purely data-driven, inductive methods for concept formation are strongly limited and modulated by one's background knowledge. It has been observed developmentally that children choose a toy monkey over a worm as being more similar to a human, but that when told that humans have spleens, they are more likely to infer that the worm has a spleen than that the toy monkey does. One conclusion is that inferences are driven by the theories that even young children have about living things and their rich internal organizations. The general claim is that concepts are formed around theories—organized systems of knowledge (Murphy & Medin, 1985). This claim is supported by evidence that concepts are more easily learned when a learner has appropriate background knowledge, suggesting that more than "brute" statistical regularities underlie our concepts.

It seems most likely that people use combinations of these representational strategies, and can deploy them flexibly based upon the categories to be learned. There is a growing consensus that at least two independent kinds of representational strategy operate in concept formation-- rule-based and similarity-based processes (Sloman, 1996). Even if one holds out hope for a unified model of all concept formation, these different representational strategies must be accommodated by that model.

Much of the most exciting current work in concept formation dispenses with the fiction that concepts are isolated, independent knowledge structures, and focuses on how concepts are related to each other. For example, the concept *hammer* seems to be inherently connected to *nails* given that one of its most commonly listed features is “used to strike nails.” In semantic networks, concepts are represented by nodes in a network, and gain their functionality by their links to other concept nodes. The meaning of a given concept is then derived from how it connects with other conceptual information.

If concepts are connected to each other, they are also importantly connected to at least two other cognitive processes: perception and language. According to this Lawrence Barsalou’s (1999) theory of Perceptual Symbol Systems, sensorimotor areas of the brain that are activated during the initial perception of an event are re-activated at a later time by association areas, serving as a representation of one’s prior perceptual experience. Because these re-activated aspects of experience may be common to a number of different events, they may be thought of as symbols, representing an entire class of events. Because they are formed around perceptual experience, they are perceptual symbols, unlike the amodal symbols typically employed in symbolic theories of cognition. Consistent with this theory, perceptual and conceptual information influence one another. For example, in cases of acquired categorical perception, people are faster at discriminating between objects that belong to different categories rather than the same category, even when those categories are arbitrarily imposed through training.

The connection between concepts and language is similarly bidirectional. In particular, one’s repertoire of concepts influences the types of word meanings that one learns, whereas the language that one speaks arguably influences the types of concepts that one forms. The former, less controversial, influence is well documented by developmental evidence suggesting that word meanings are relatively easily acquired to the extent that the words map onto existing concepts that capture the correlational structure of the world. There is growing evidence amassing for the latter influence (Gentner & Goldin-Meadow, 2003) that demonstrates that words can selectively direct attention to particular dimensions, facilitate certain thought patterns (by increasing their likelihood

of being triggered or creating tokens for complex ideas), and structure otherwise chaotic cognitive domains. Overall, the connections between concepts and perception on the one hand and between concepts and language on the other hand reveal an important dual nature of concepts. Concepts are used both to recognize objects and to ground word meanings. Working out the details of this dual nature will go a long way towards understanding how human thinking can be both concrete and symbolic.

## References

- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. Psychological Review, 93, 154-179.
- Barsalou, L.W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577-660.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). A study of thinking. New York: Wiley.
- Fodor, J. A. (1998). Concepts: Where cognitive science went wrong. Oxford: Oxford University Press.
- Gentner, D., & Goldin-Meadow, S. (Editors) (2003). Language in mind: Advances in the study of language and thought. Cambridge, MA: MIT Press.
- Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. Psychological Review, 92, 289-316.
- Nosofsky, R. M (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115, 39-57.
- Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-232.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119, 3-22.

### **Supplemental Readings**

Estes, W. K. (1994). Classification and cognition. New York: Oxford University Press.

Margolis, E., & Laurence, S. (Editors) (1999). Concepts: Core readings. Cambridge, MA: MIT Press.

Murphy, G. L. (2002). The big book of concepts. Cambridge, MA: MIT Press.